

KOREAN PATENT ABSTRACTS

(11)Publication number: **10-1999-0013736 A**
(43)Date of publication of application: **25.02.1999**

(51)Int. Cl.
G06K 5/00

(21)Application number: **10-1998-0027734**
(22)Date of filing: **02.07.1998**
(30)Priority: **97-176822 (02.07.1997) JP**
(71)Applicant: **MATSUSHITA ELECTRIC IND
CO LTD**
(72)Inventor: **SATO MITSUHIRO
NOGUCHI NAHIKO
SUGANO YUJI
NOMOTO MASAKO
INABA MITSUAKI
FUKUSHIGE TAKAO**

**(54) AUTOMATIC EXTRACTION DEVICE FOR RELATIVE KEYWORD,
DOCUMENT RETRIEVING DEVICE AND DOCUMENT RETRIEVING
SYSTEM USING THESE DEVICES**

(57) Abstract:

PURPOSE: To automatically extract a relative keyword which is matched with the characteristics of a document to be practically retrieved and which is capable of obtaining one or more retrieval results at the time of executing retrieval using the keyword.

CONSTITUTION: An automatic extraction device for relative keywords is provided with a document set selection part 19 for specifying a partial set of each document based on the attribute information, input retrieval expression, etc., of the document, a word statistic information management part 17 for managing the statistic information of respective words in the whole objective document 11 and words appearing in each document as well as their statistic information 15; and a word ranking part 18 for calculating the importance of each word appearing in a partial set of a certain document and for aligning respective words in the order of importance, wherein the management part 17 quickly finds out the statistic information of respective words in the whole document and a specified partial set of the document. Consequently, words appearing in a certain document set can be ranked based on their importance and a part of the ranked words can be presented as a relative keyword.

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(51) Int. Cl.⁸
G06K 5/00

(11) 공개번호 특1999-013736
(43) 공개일자 1999년02월25일

(21) 출원번호 특1998-027734
(22) 출원일자 1998년07월02일
(30) 우선권주장 97-176822 1997년07월02일 일본(JP)
(71) 출원인 마쓰시타덴키산교(주) 모리시타요오이치
일본 오사카부 가도마시 오오야자 가도마 1006
(72) 발명자 사토미츠히로
일본 가나가와켄 아츠기시 미야노사토 4-11-11
노구치나오히코
일본 요코하마시 고호쿠쿠 후토오조 1323-601
간노유지
일본 요코하마시 미나미쿠 오오카 1-1-19-504
노모토마사코
일본 사이타마켄 도코로자와시 미나미 스미요시 7-18-504
이나바미츠아키
일본 도쿄도 오오타쿠 히가시 로쿠고 2-20-5-628
추쿠시게요시오
일본 가나가와켄 후지사와시 다이기리 3-4-2-비-303
(74) 대리인 이병호, 최달용

심사청구 : 있음

(54) 키워드 추출 시스템 및 그를 사용한 문서 검색 시스템

요약

텍스트 검색 시스템에서 탐색을 용이하게 하기 위해 키워드를 제공하는 시스템에 관한 것이다. 텍스트 베이스(text base)를 구성하는 각 텍스트에 대해, 상기 시스템은 텍스트에 사용된 각 단어의 단어 ID와 대응 단어의 단어 발생 카운트를 만든다. 단어 발생 카운트는 각 텍스트내의 단어의 발생수를 나타낸다. 텍스트 베이스를 구성하는 텍스트에 사용된 각 단어에 대해, 상기 시스템은 총 단어 발생 카운트와, 단어를 포함하는 텍스트의 수를 나타내는 포함 텍스트 카운트를 만든다. 선택된 텍스트에 포함된 각 단어에 대해, 단어 발생 카운트와 총 단어 발생 카운트와 포함 텍스트 카운트를 사용하여 중요도를 계산한다. 선택된 텍스트에 포함된 단어는 중요도순으로 소트된다. 적어도 소트된 단어들중 일부는 관련 키워드로서 디스플레이 된다.

대표도

도1

명세서

도면의 간단한 설명

도 1은 본 발명이 구현될 수 있는 컴퓨터의 전형적인 구성을 도시하는 개략 블록도.

도 2는 관련 키워드 추출 시스템(100)이 본 발명에 따라 내장되는 문서 검색 시스템(2)의 소프트웨어의 전형적인 구성을 도시한 개략도.

도 3은 문서 속성표의 전형적인 기록 구조를 도시하는 도면.

도 4는 도 2의 블록 정보 프로세서에 의해 만들어지고 유지된 사전 및 표를 도시하는 도면.

도 5는 새로운 문서 텍스트가 DB(70)에 추가되는 경우 사전(210)과 표(220, 230)를 유지하기 위해 실시되는 프로세서(200)의 제1 루틴의 동작을 도시하는 흐름도.

도 6은 문서(T1)가 DB(70)로부터 삭제되는 경우 표(220, 230)를 유지하기 위해 실시되는 프로세서(200)의 제2 루틴의 동작을 도시하는 흐름도.

도 7은 이용자와 대화식으로 사용하는 문서 검색 시스템(2)의 동작 흐름을 도시하는 흐름도.